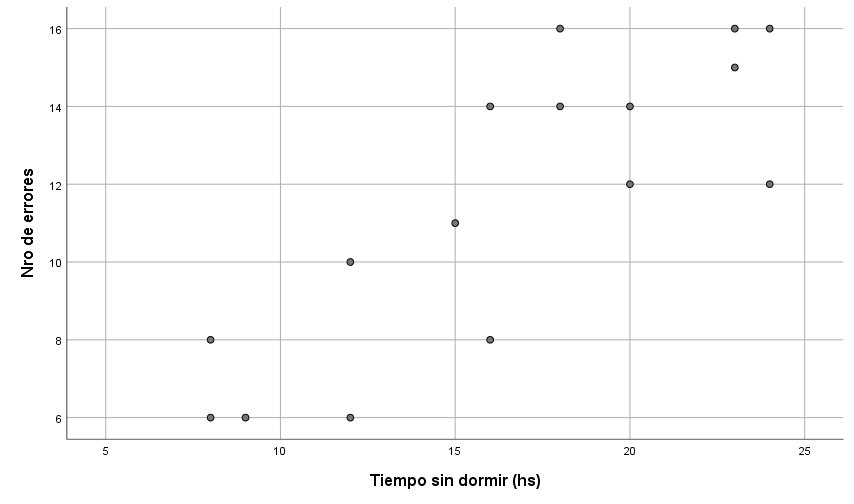
a) Grafica de dispersión

A simple vista no parece haber una relación lineal entre las variables, los puntos están muy dispersos entre sí, aunque se puede apreciar una cierta tendencia a aumentar los errores a medida que aumentan las horas sin dormir.

b) Modelo de regresión lineal:

Se busca la función de x más simple, lineal, que permita aproximar el valor de Y, mediante la fórmula:

X: Variable independiente, predictora o explicatora, generalmente un dato en base al cual se quiere encontrar su valor correspondiente de la variable y.

Y: Variable dependiente, predicha o explicada, la cual varía en función de x.

Generalmente el valor de Y no coincide con el valor de , ya que se encuentra afectada por un error el cual se denomina residuo o error residual:

La recta muestral que estima a la población será:

Donde es la ordenada al origen, la pendiente, y es una variable aleatoria con E()=0 y Var()=.

El supuesto del modelo es que los residuos son independientes y se distribuyen en forma normal con media 0 y varianza constante , de no cumplirse esto, el modelo será invalido.

c) Siendo las variables:

X: Tiempo sin dormir(hs)

Y: Nro de errores

Se obtiene la siguiente tabla de Excel:



La ecuación de la recta finalmente será:

d) Intervalo de 99% de confianza para β.

e) Para poder estimar la cantidad de errores en un sujeto con 18,5 horas sin dormir, solo basta con reemplazar x=18,5 en la ecuación de la recta

Como el número de errores es un numero entero, se estima que cometerá 13 errores.

f) No, el modelo es utilizable solo para valores que se encuentren entre el valor máximo y mínimo de los datos usados para construirlo.

g) El coeficiente de correlación muestral es el denominado r el cual se calcula de la siguiente manera:

Este determina la relación lineal entre las variables de estudio.

h)

i) El coeficiente de determinación muestral es simplemente r2=0,69, este explica el porcentaje de la dependencia de la variable dependiente en función de la independiente, para este caso, 69%.

j) Test:

H0: 𝜌=0; H1: 𝜌>0

El p-value de ese valor t es un número muy pequeño, menor a 0,01

Como el p-value obtenido es menor al nivel de significancia se rechaza la hipótesis nula y se puede afirmar que 𝜌 es mayor a 0.

k) Test:

H0: Los residuos se distribuyen de manera normal

H1: Los residuos NO se distribuyen de manera normal

Salida del software:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Pruebas de normalidad** | | | | | | |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Estadístico | gl | Sig. | Estadístico | gl | Sig. |
| Unstandardized Residual | ,139 | 16 | ,200\* | ,965 | 16 | ,758 |
| \*. Esto es un límite inferior de la significación verdadera. | | | | | | |
| a. Corrección de significación de Lilliefors | | | | | | |

Como n<50 utilizamos el test de Shapiro-Wilk, se observa que el p-value calculado es 0,758, un número muy elevado, por el cual no rechazamos H0 y podemos decir nuestro modelo es válido.